## § 1.4. Нечеткие системы кластеризации

Взаимосвязь между кластерным анализом и теорией нечетких множеств основана на том обстоятельстве, что при решении задач структуризации сложных систем большинство формируемых классов объектов размыты по своей природе. Эта размытость состоит в том, что переход от принадлежности к непринадлежности элементов к данным классам скорее постепенен, чем скачкообразен. Требуется ответ не на вопрос «Принадлежит ли рассматриваемый элемент тому или иному классу или нет?», а на вопрос: «В какой степени данный элемент принадлежит рассматриваемому классу?»[14].

В общем случае задачей нечеткой кластеризации является нахождение нечеткого разбиения или нечеткого покрытия множества элементов исследуемой совокупности, которые образуют структуру нечетких кластеров, присутствующих в рассматриваемых данных. Эта задача сводится к нахождению степеней принадлежности элементов искомым нечетким кластерам, которые в совокупности и определяют нечеткое разбиение или нечеткое покрытие исходного множества рассматриваемых элементов.

**Общая формальная постановка задачи нечеткого кластерного анализа.** Пусть исследуемая совокупность данных представляет собой конечное множество элементов , которое получило название множество объектов кластеризации. В рассмотрение также вводится конечное множество признаков , каждый из которых количественно представляет некоторое свойство или характеристику элементов рассматриваемой проблемной области. При этом –общее количество объектов, а – количество признаков объектов [14].

Предполагается, что для каждого из объектов кластеризации измерены все признаки множества в некоторой количественной шкале. Каждому из элементов поставлен в соответствие вектор , где – количественное значение признака для объекта данных . Для определенности будем полагать, что все принимают некоторые действительные значения, т.е. .

Процесс измерения свойств может быть реализован в различных шкалах, каждая из которых характеризуется допустимым преобразованием данных. Само множество признаков следует выбирать таким образом, чтобы все признаки были измерены в шкале отношений или шкале интервалов. В этом случае результаты нечеткой кластеризации имеют содержательную интерпретацию, адекватную проблеме нахождения нечетких кластеров [14].

Векторы удобно представить в виде матрицы данных размерности , каждая строка которой равна значению вектора.

Задача нечеткого кластерного анализа формулируется следующим образом: на основе исходных данных D определить такое нечеткое разбиение множества на заданное число c нечетких кластеров , которое доставляет экстремум некоторой целевой функции среди всех нечетких разбиений.

**Уточненная постановка задачи нечеткой кластеризации.** Один из вариантов конкретизации задачи нечеткого кластерного анализа основан на алгоритме ее решения методом нечетких -средних.

Дадим определение нечеткому покрытию нечеткого множества. Система нечетких подмножеств нечеткого множества называется нечетким покрытием, если выполняется следующее условие:

(1)

т.е. объединение всех подмножеств из совпадает (или «покрывает») с исходным нечетким множеством .

Для уточнения вида целевой функции в рассмотрение вводятся некоторые дополнительные понятия. Предполагается, что искомые нечеткие кластеры представляют собой нечеткие множества , образующие нечеткое покрытие исходного множества , при этом условие (1) принимает вид:

(2)

где – общее количество нечетких кластеров , которое считается предварительно заданным и [14].

Далее для каждого нечеткого кластера вводятся в рассмотрение центры искомых нечетких кластеров , которые рассчитываются для каждого из нечетких кластеров по следующей формуле:

 (3)

где m-некоторый параметр, называемый экспоненциальным весом и равный некоторому действительному числу (m > 1). Каждый из центров кластеров представляет собой вектор  в некотором q-мерном нормированном пространстве [14].

В качестве целевой функции будем рассматривать сумму квадратов взвешенных отклонений координат объектов кластеризации от центров искомых нечетких кластеров:

 (4)

где m-экспоненциальный вес нечеткой кластеризации , значение которого задается в зависимости от количества элементов множества . Чем больше элементов содержит множество , тем меньшее значение выбирается для .

Задача нечеткой кластеризации может быть сформулирована следующим образом: для заданных матрицы данных , количества нечетких кластеров и , параметра определить матрицу значений функций принадлежности объектов кластеризации нечетким кластерам , которые доставляют минимум целевой функции (4) и удовлетворяют ограничениям (2), (3), а также дополнительным ограничениям (5) и (6):

 (5)

 (6)

Условие (5) исключает появление пустых нечетких кластеров в искомой нечеткой кластеризации. Последнее условие (6) имеет формальный характер, поскольку непосредственно следует из определения функции принадлежности нечетких множеств. В этом случае минимизация целевой функции (4) минимизирует отклонение всех объектов кластеризации от центров нечетких кластеров пропорционально значениям функций принадлежности этих объектов соответствующим нечетким кластерам [14].

Недостатком постановки задачи нечеткой кластеризации является необходимость априорного задания общего числа нечетких кластеров.